LINKING DATA TO SAVE LIVES

Case study of the Massachusetts Public Health Data Warehouse and the use of data to address opioid overdoses

Jane Wiseman

Institute for Excellence in Government, January 2024

Linking data saves lives: Case study of the Massachusetts Public Health Data Warehouse and the use of data to address opioid overdoses

Jane Wiseman, Institute for Excellence in Government

January 2024

Connecting data across silos is challenging. Yet in government, the most complex and important problems often involve multiple agencies – for example homelessness is seldom just a housing problem but rather one that touches on employment, health, mental health, family dynamics and sometimes incarceration or substance addiction. Solving problems by connecting data sets about people but in a secure way that protects their privacy is similarly challenging but important. This case study describes an early success in linking data in Massachusetts that remains one of the best examples in the country of secure person-level data linkage across silos that fuels policy insight while providing appropriate privacy and security protections.

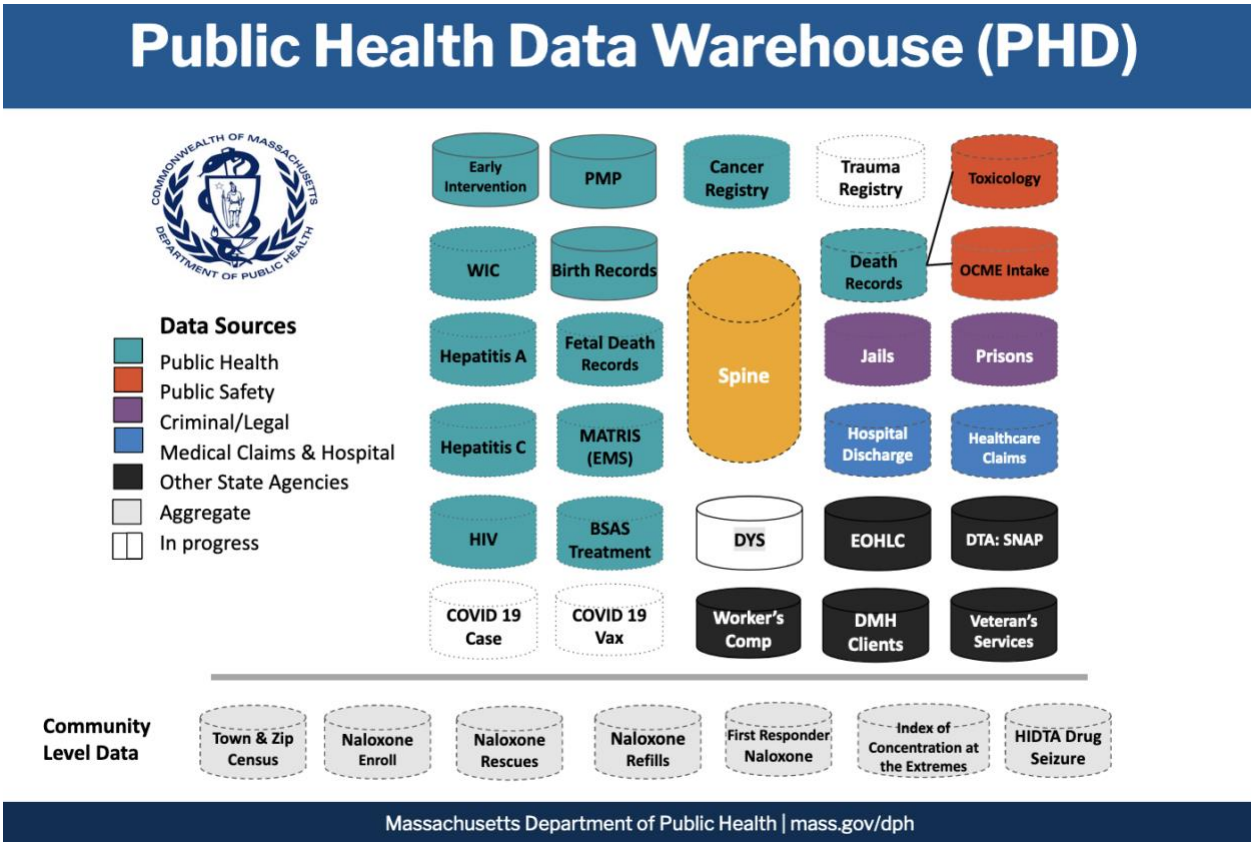
At the federal level, as the idea for a National Secure Data Service is tested via pilot and demonstration projects, here in Massachusetts the ability to link data in a secure way has been proven by the Massachusetts Public Health Data Warehouse (PHD). This unique model links person-level data on the fly, behind a firewall so that insights can be garnered without compromising the identity of individuals in the dataset. The sophisticated analytics in this model are driven by a governance structure that puts privacy first. Researchers and policymakers can feed queries into the data warehouse but are never provided with any individual-level data, only the aggregated de-identified results of their queries.

The Massachusetts PHD links person-level data across 38 distinct sources including vital records, public health, healthcare, social services, and justice to provide near real-time insight on emerging threats to public health, such as drug overdose fatalities and near fatalities. The PHD provides secure access for authorized researchers to this individually-linked, longitudinal, and privacy-protected administrative data.

The PHD was created in 2017 to link data across government to uncover insights on the opioid epidemic that were not available in any single system alone. Since then, the PHD has grown to address other emerging public health issues, such as maternal and child health disparities, substance use, COVID-19, and the effects of climate change on health.

The magnitude of coverage of the data, with on average, 93% of the state's population included, creates a powerful platform for research, eliminating the small sample sizes or self-selection bias that may be inherent in health studies that voluntarily enroll participants, and reducing the attrition that often occurs in a voluntary participation research cohort over the lifecourse.

The PHD links mental health data, jail and prison data, vital records, substance addiction treatment data, ambulance encounter information, the state's all-payer health claims insurance database, and others. The chart that follows shows how the many data sources came together:



The ability to conduct analysis on individual-level data from disparate data sets while protecting data privacy required innovation. Individual level data are temporarily linked and never stored in a common database, going back to their original files once the analysis is complete. Data are only temporarily linked, and calculations are completed without the source data being available for download. High-level encryption assures that data transfers are secure.

The result satisfies and exceeds all federal requirements, including HIPPA privacy requirements as well as the federal regulatory (42CFR part 2) requirements that provide additional protections to data about individuals receiving substance abuse services. Forging an innovative way to protect privacy while doing individual-level data matching took patience in working across disciplines with countless multi-hour meetings to hammer out all the details. The chart that follows highlights some of the key methods to preserve privacy while linking data.

Privacy preserving data linkage in the PHD
PHD security measures to protect data include: • Encrypted identifiers from each dataset are linked to encrypted identifiers from the All Payers Claims Database (APCD), which serves as the spine for linking individual information across datasets, and assigned a random project specific ID • Linkage is made using six personal identifiers, which have been encrypted: first name, last name, date of birth, sex, social security number, and zip code of residence. • After each dataset is linked, the encrypted personal identifiers (except for the zip codes and sex) are dropped to protect the information. • Dates of service included in some datasets, such as hospitalization data, are masked so that this information is retained for analytic purposes while still maintaining the privacy of the data.
Source: Adapted from Massachusetts Department of Public Health, Legislative Report, 2021. https://www.mass.gov/lists/public-health-data-warehouse-phd-publications#legislative-reports-

Early successes of the PHD include insights only available by combining data sets that had never before been linked. For example, linking data showed that the rate of opioid use disorder (OUD) in Massachusetts were nearly four times higher than national estimates and two times higher than what past surveillance methods would have found for the state. Also, linking individual data across systems identified that nearly 10% of non-fatal overdoses were followed by a fatal overdose within two years. This insight led to policy changes that enabled earlier intervention with individuals who had experienced non-fatal overdoses. Another early insight was that few opioid overdose patients were prescribed methadone or buprenorphine when they were brought to the emergency room, but for those who were prescribed the treatment drugs, lives were saved. This insight resulted in the state passing a law mandating that non-fatal overdose patients be offered the option to start these medications before being discharged from the hospital after their emergency room visit.

Because the PHD linked multiple datasets, DPH was able to identify more populations at risk for opioid-related overdose deaths than possible with typical surveillance methods. Important early findings from the PHD, regarding opioid overdoses (covering the initial inquiry period, with data from 2011 – 2015) include:

- Those recently released from incarceration had an opioid overdose rate 120 times higher than the rest of the adult population.
- Individuals experiencing homelessness where 64% more likely to have a fatal opioid overdose and nearly 6 times more likely to experience a non-fatal overdose compared to the rest of the adult population.
- Individuals diagnosed with a serious mental illness (mood disorders, schizophrenia, and other psychotic disorders) had six times the risk of fatal opioid overdose compared to the rest of the adult population.

These data insights produced policy change, including treatment availability at county jails and proactive approaches with individuals experiencing homelessness. Further, prior to the development of the PHD, overdose death data had taken as much as two years to be finalized due to backlogs at the Office of the Chief Medical Officer for drug toxicology reports. This led DPH to create a predictive model to estimate opioid overdose deaths based on preliminary data, which significantly closed the gap between opioid deaths and actionable insight to prevent them.

No one agency had all of the necessary skills or resources, nor sufficient manpower for interagency data sharing. The most effective teams leverage diverse talents in new ways. For example, in Massachusetts the opioid data initiative required legal, data, technology, and analytic skills that were found in the Department of Public Health, the Center for Health Information and Analysis, the Executive Office of Technology Services and Security, along with data, legal and privacy experts from across the many agencies that shared data.

The PHD is not solely a government or academic pursuit, but rather engages with community stakeholders to assure that input is heard on priorities for research and action. The Community Advisory Board (CAB) was established to provide community voices and perspectives to equitably guide data collection and analysis. For example, the CAB helps define the analytic questions to be answered via research and provides input on dissemination of research results.

With enhanced computing power and additional datasets added over the years since it was created, the PHD remains an important resource as Massachusetts looks to address its public health priorities and disease prevention.

Acknowledgements: The author would like to thank Katie O'Toole of the Data Foundation for editorial support, and Dana Bernson, Director, Special Analytic Projects, Office of Population Health, Massachusetts Department of Public Health for proving accurate data as of the date of publication of this case study. All errors and omissions are the sole responsibility of the author.